Model-independent anomaly detection in gravitational waves

Emilie Hertig[1], supervised by Inar Timiryasov[2] and Sergey Sibiryakov[3]

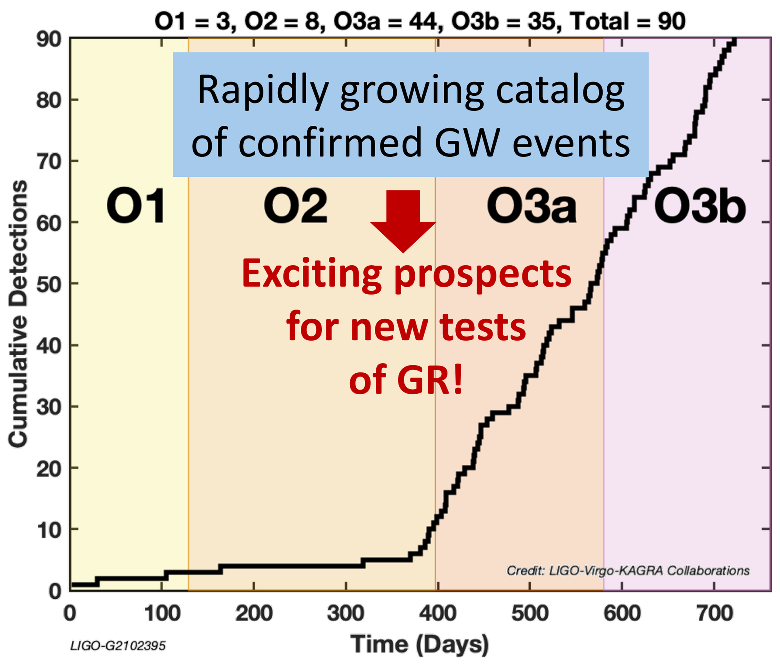
Based on master's thesis completed at EPFL (Lausanne, Switzerland)

[1]Institute of Astronomy, University of Cambridge; [2]Niels Bohr Institute, University of Copenhagen; [3]Perimeter Institute for Theoretical Physics

1. Introduction

Probing strong gravity

Tensions in cosmology, Hawking's black hole **information paradox** and the absence of a full **quantum description of General Relativity** (GR) point towards an **incomplete understanding of gravity**. This motivates the search for **deviations from GR**, which are most likely to be found in the **strong and dynamical regime** probed by **binary black hole mergers**. New physics may then manifest as **anomalies** in the resulting spacetime metric perturbations known as **gravitational waves (GW)**.



Model independence

Inferred parameters such as GW source position are **highly uncertain** (e.g. the best localized event so far has 10^3-10^4 galaxies in its 90% credible volume!), and a **wide variety of possible extensions of GR** exist. Tests based on **prior knowledge** of anomalous waveform features are therefore **suboptimal**.

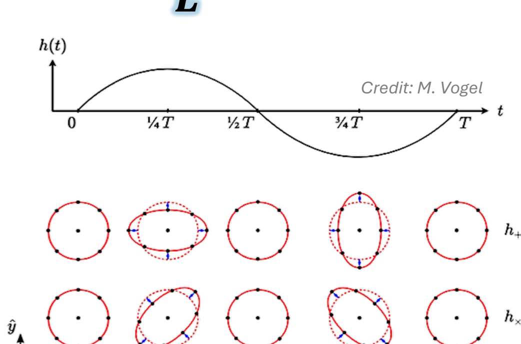
➔ Goal of this work: use an artificial **neural network** as a **flexible fitting tool** to detect GW anomalies **without restrictive assumptions** of the underlying theory.

2. GW detection

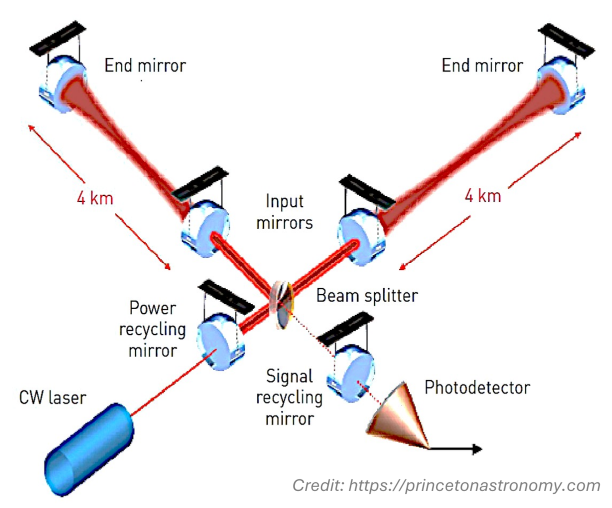
Strain from passing GW

- Alternating **distortions** along polarization axes

$$\frac{\Delta L}{L} \sim 10^{-21}$$



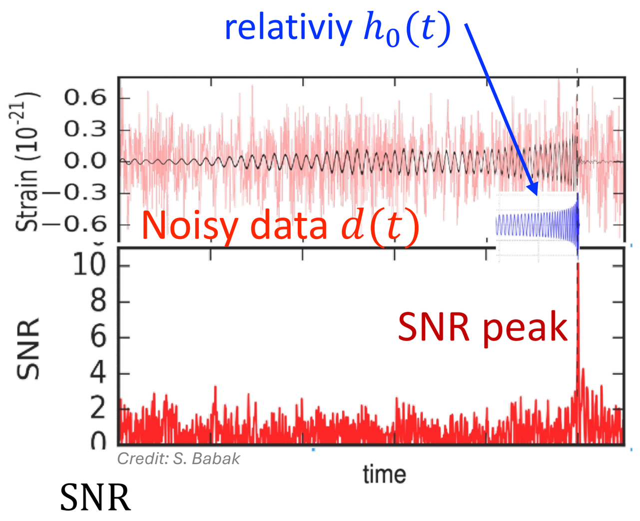
$L_{\text{eff}} \sim 1200$ km
$\Delta L \sim 10^{-15}$ m, size of a proton!



Interferometer, e.g. LIGO

Whitening, **matched filtering**

GW template from numerical relativiy $h_0(t)$



$$\text{SNR} = \frac{1}{\sigma_h^2} \Re \int_0^\infty \tilde{d}^*(f) \tilde{h}_0(f) e^{-2\pi i f t} df$$

3. Neural network

Data

- Training set is **1 GW event**
- Goal is to **overfit the training set** => no test set
- Time series sampled at 4096 Hz

Activation

- Fast oscillations best fitted with **sinusoidal** activation

Architecture

- Multilayer perceptron (**MLP**), 2 hidden layers of **6 neurons**
- As **small** as possible: **complexity penalized** in model selection

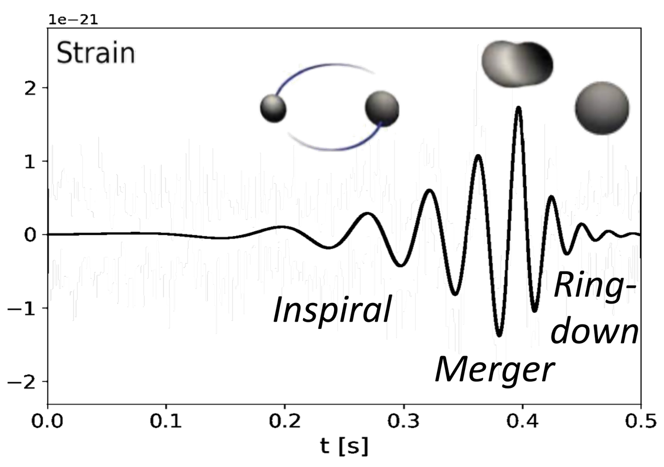
Loss function

Data & model are **whitened**

$$-\log \mathcal{L}(\mathbf{d}|\mathbf{h}) = 2\Re \int_0^\infty (\tilde{d}(f) - \tilde{h}(f))(\tilde{d}^*(f) - \tilde{h}^*(f)) df$$

4. Methodology

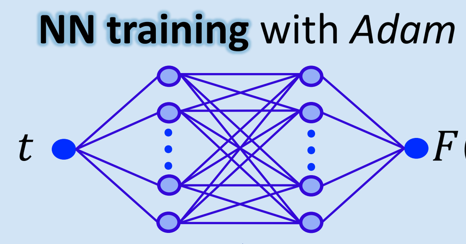
Smooth analytical **toy model** qualitatively mimicking GW signal



Add **anomalous feature**
- Localized glitches
- Physically-motivated waveforms (Sec. 5)

Inject into **2 independent** Gaussian white **noise realizations**
➔ Mock whitened LIGO dataset

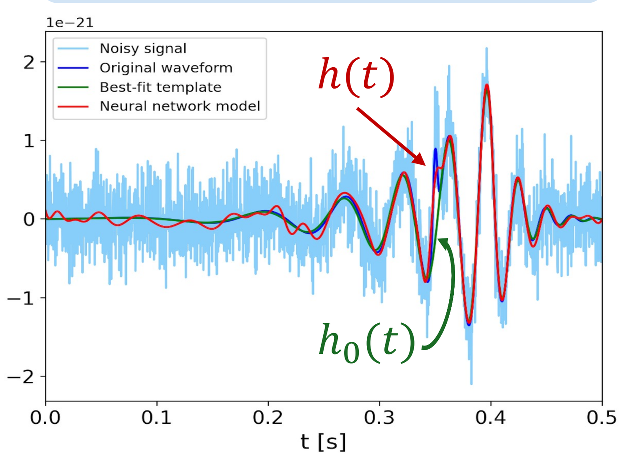
NN training with *Adam*



Differential evolution
- Find best-fit **base model template** $h_0(t)$
- 6 free parameters for our toy model

Alternative hypothesis

$$h(t) = h_0(t) + F(t)$$



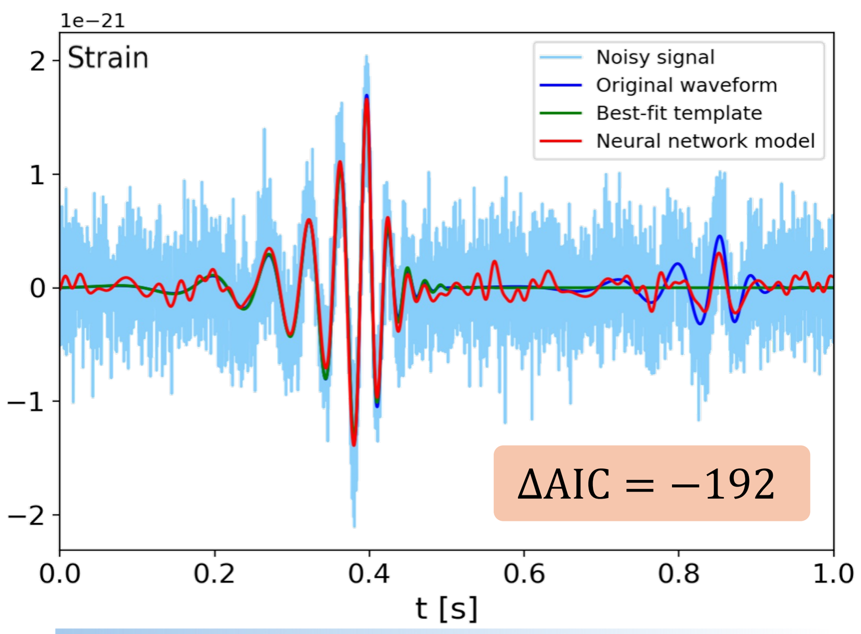
Model selection: **Akaike Information Criterion (AIC)**

$$\text{AIC} = -2\log\mathcal{L}(\mathbf{d}|\mathbf{h}) + 2k$$

Number of model parameters

- h preferred if $\textbf{AIC}(h) - \textbf{AIC}(h_0) \leq -10$
- Estimates **loss of information**
- Selects **closest fit** to data **without assuming existence** of true model

5. Results

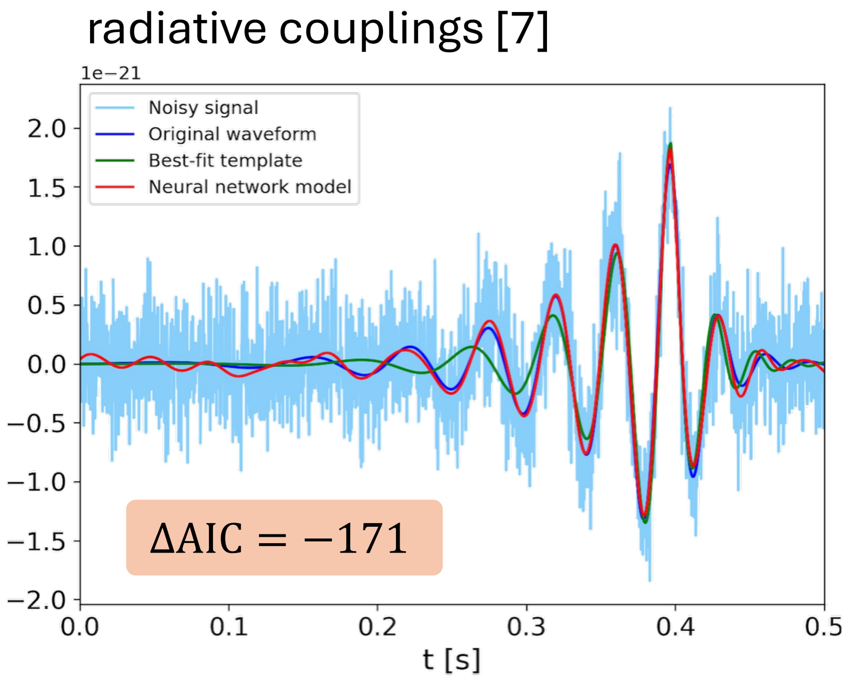


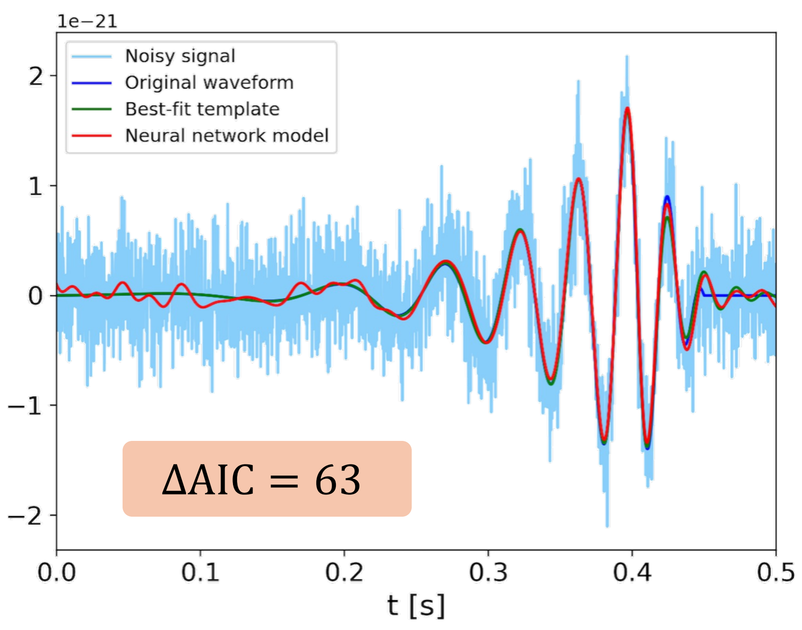
Echo (A)

Quantum structure at BH horizon (e.g. firewall) producing **damped reflections** of emitted GW [6]

Inspiral dephasing (B)

EFT of gravity: higher order corrections to **BH potentials** and radiative couplings [7]





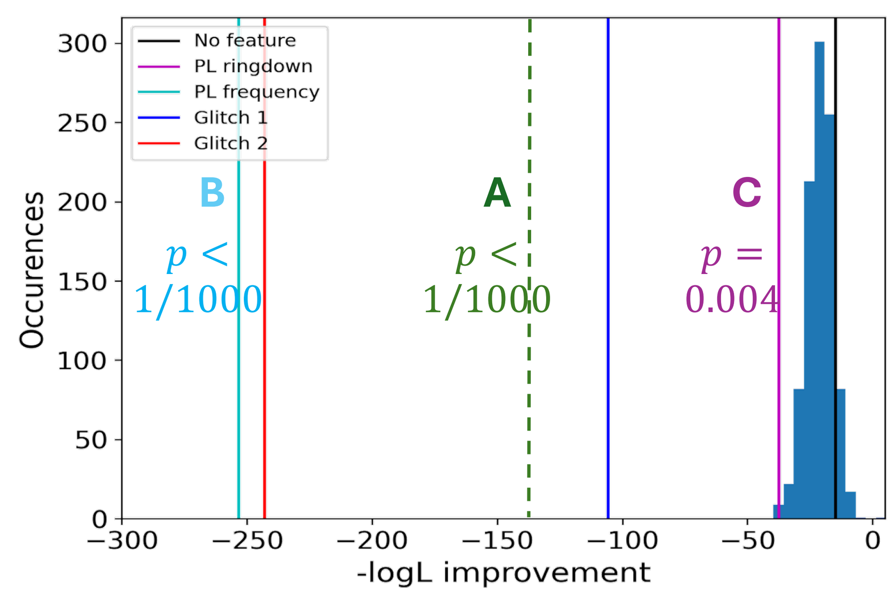
Ringdown modulation (C)

Different **quasinormal modes** of **exotic compact objects** [8] / BH in **modified gravity** theories [9]

Monte Carlo analysis: computing ***p*-values**

- Compare $\Delta(-\log\mathcal{L})$ for anomalous signals & base model in 1000 indep. noise realizations

Detections confirmed => successful proof of concept on simplified mock data!

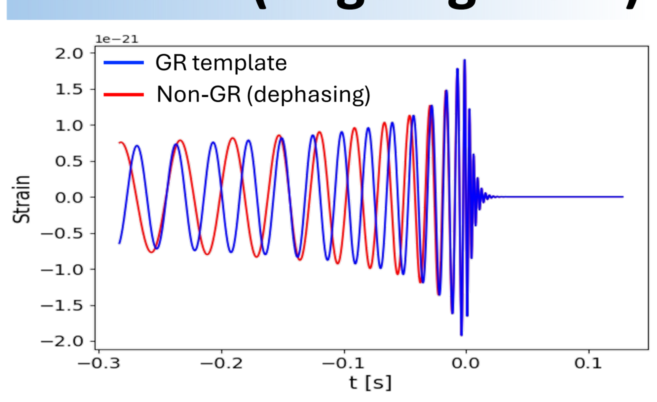


6. Towards application to real data

Challenges

- Full **GR waveforms** are **more complex** than toy model
- Quantitatively **realistic features** may be **fainter**
- Detector output is not just data + noise: account for **2 polarizations** and **antenna response functions**

Solution (ongoing work)



- Implemented **GR waveforms**, realistic features and **detector injection** in **Bilby**
- Differentiable likelihood in **Pytorch**
- Working on **schedule-free training** [10] and **systematic NN architecture search**

Selected references

[1] B. P. Abbott *et al.*, "Observation of Gravitational Waves from a Binary Black Hole Merger" (2016), arXiv: 1602.03837.
[2] J. Aasi *et al.*, "Advanced LIGO" (2015), arXiv: 1411.4547.
[3] R. T. D'Agnolo and A. Wulzer, "Learning New Physics from a Machine" (2019), arXiv: 1806.02350.
[4] Y. Yang, "Can the strengths of AIC and BIC be shared? [...]", *Biometrika* **92** (2005).
[5] The LIGO and Virgo Collaborations, "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run" (2023), arXiv:2111.03606.
[6] J. Abedi *et al.*, "Echoes from the Abyss: Tentative evidence for Planck-scale structure at black hole horizons" (2016), arXiv:1612.00266.
[7] S. Endlich *et al.*, "An effective formalism for testing extensions to General Relativity with gravitational waves" (2017), arXiv:1704.01590.
[8] A. Toubiana *et al.*, "Modeling gravitational waves from exotic compact objects" (2020), arXiv:2011.12122.
[9] L. Manfredi *et al.*, "Quasinormal Modes of Modified Gravity (MOG) Black Holes" (2017), arXiv:1711.03199.
[10] A. Defazio *et al.*, "The Road Less Scheduled" (2024), arXiv:2405.15682.

emh83@cam.ac.uk